Urinary 1H-NMR metabolomic fingerprinting reveals biomarkers of pulse consumption related to energy-metabolism modulation in a subcohort from the PREDIMED study

Francisco Madrid-Gambin†, Rafael Llorach*,†, Rosa Vázquez-Fresno†, Mireia Urpi-Sarda†, Enrique Almanza-Aguilera†, Mar García-Aloy†, Ramon Estruch§#, Dolores Corella-⊥ and Cristina Andres-Lacueva*†

†Biomarkers & Nutrimetabolomics Laboratory. Nutrition, Food Science and Gastronomy Department, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona, Spain.

§Department of Internal Medicine, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona, Spain.

⊥Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain.

#CIBER OBN, The Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid, Spain.

*E-mail: rafallorach@ub.edu. Tel: +34.934033798. *E-mail: candres@ub.edu. Tel: +34.934034840. Fax: +34.93403593.
Abstract

Little is known about the metabolome fingerprint of pulse consumption. The study of robust and accurate biomarkers for pulse dietary assessment has great value for nutritional epidemiology regarding health benefits and their mechanisms. To characterize the fingerprinting of dietary pulses (chickpeas, lentils and beans), spot urine samples from a subcohort from the PREDIMED study were stratified, using a validated food frequency questionnaire. Non-pulse consumers (≤ 4 g/day of pulse intake) and habitual pulse consumers (≥ 25 g/day of pulse intake) were analysed using a 1H-NMR metabolomics approach combined with multi- and univariate data analysis. Pulse consumption showed differences through 16 metabolites coming from (i) choline metabolism, (ii) protein-related compounds, and (iii) energy metabolism (including lower urinary glucose). Stepwise logistic regression analysis was applied to design a combined model of pulse exposure, which resulted in glutamine, dimethylamine and 3-methylhistidine. This model was evaluated by receiver operating characteristic curve (AUC > 90% in both training and validation sets). The application of NMR-based metabolomics to pulse exposure highlighted new candidates for biomarkers of pulse consumption, the role of choline metabolism and the impact on energy metabolism, generating new hypotheses on energy modulation. Further intervention studies will confirm these findings.

Keywords

pulses, legumes, metabolomics, NMR, choline metabolism, energy, biomarkers, ROC curve
The Mediterranean diet (MD) is a dietary pattern characterized by a high intake of vegetables, cereals, pulses, nuts, fish and olive oil, low intake of red meat and processed meat products, and low to moderate consumption of poultry, wine and dairy products.\(^1\) Moreover, the MD has been demonstrated to be useful in the prevention of type 2 diabetes, obesity, inflammatory diseases, cardiovascular diseases (CVD) and even cancer.\(^2\text{-}^5\)

One of the components of the MD is pulses, which constitute an excellent food, providing protein, dietary fibre, many vitamins and minerals, as well as a great variety of phytochemicals.\(^6\text{-}^8\) Thus, they could contribute to the beneficial effects reported for this dietary pattern.\(^9\) In addition, pulses are increasingly being recognized for their role in promoting good health.\(^6,10\text{-}^12\) Indeed, habitual pulse consumption is included in the main dietary guidelines worldwide, including the MD,\(^13\) the Dietary Guidelines for Americans\(^14,15\) and the Nordic Diet,\(^16\) among others, and they are also advocated in view of their low environmental impact compared with other protein sources.\(^17\)

Metabolomics is a powerful tool for identifying food exposure biomarkers in humans\(^18\) and provides new information on dietary components and dietary patterns.\(^19\) In this regard, the evaluation of dietary exposure through a combination of biomarkers enables a better understanding of compliance to a dietary exposure.\(^20\) Moreover, little is known about the metabolome fingerprint from legume consumption either individually or as a complex food group, with only a few tentative biomarkers being described.\(^21,22\)
Determining the changes in the urinary metabolome, new biomarkers of intake and/or their effect may reveal potential modifications in diet-related physiology both in healthy and diseased individuals. Furthermore, metabolomic approaches have been proposed for evaluating the relationship between nutrition and health status. In light of this connection, recent scientific publications have pointed out the potential health benefits of legumes in chronic diet-related diseases, such as CVD and type 2 diabetes mellitus. Thus the application of nutrimetabolomics to a high-cardiovascular-risk population could provide new insights into this potential relationship.

In the present work, we compared the metabolome profiles of reported pulse consumption in a free-living population to find putative biomarkers reflecting intake and/or effect of intake. Analysis of individuals under free-living conditions enables more representative data to be obtained on the metabolome fingerprints of pulse consumers. In light of this, a better understanding of the specific role of pulse consumption in terms of health benefits, beyond their excellent nutritional profile, is expected. Therefore, the aim of the present study was to investigate dietary pulse fingerprinting in spot urine using an untargeted $^1$H-NMR metabolomic approach on a free-living subcohort from the PREDIMED study. For this purpose, we mainly focused on urinary biomarkers of a complex pulse exposure comprising chickpeas, lentils and beans in a combined urinary biomarker model.

2. Material and methods

2.1. PREDIMED subcohort study
For the present study, a subsample of 50 participants from the PREDIMED study (ISRCTN 35739639; http://www.predimed.org) was taken. The PREDIMED study is a large, parallel-group, multicentre, randomized and controlled clinical trial assessing the effects of an MD on the primary prevention of CVD. The trial protocol was conducted according to the Declaration of Helsinki and was approved by the Institutional Review Boards of all the centres involved. Briefly, free-living participants (55–80 years old) without CVD that fulfilled at least one of the two following criteria – type 2 diabetes mellitus or three or more major cardiovascular risk factors – were included for an MD supplemented either with extra virgin olive oil or mixed nuts. The exclusion criteria were CVD, any severe chronic illness, drug or alcohol addiction, a history of allergy, or intolerance to olive oil or nuts. The subcohort consisted of a random sample of participants at high cardiovascular risk, recruited from the Barcelona and Valencia PREDIMED centres. The PREDIMED study design and 137-item validated food frequency questionnaires (FFQs) used have been reported elsewhere. Data reported from the FFQs included information on total legume consumption, and disaggregated type of legume consumed.

2.2. Stratification of the study population

2.2.1. Defining potential consumers

Both the use of FFQs and the population stratification of a cohort of individuals by consumption have demonstrated an effective approach for the study of biomarkers of food consumption. Participants were classified into two levels (consumers and non-consumers) of habitual intake of dietary pulse foods.
(chickpeas, lentils or beans) based on the analysis of the validated FFQs 
(Supporting Information, Table S1). Intake of pulses was calculated as the sum 
of consumed chickpeas, lentils and beans. Non-pulse (NP) consumers were 
defined as subjects with sporadic or non-consumption (≤ 4.00 g/day) of pulses. 
Habitual pulse (HP) consumers were set at a consumption of ≥ 25.71 g/day, 
regularly. In order to explore global pulse consumption, individuals that did not 
consume the three kinds of pulses simultaneously were also excluded. 
Additionally, the condition of sporadic or non-intake of peas (≤ 4 g/day) was 
taken into consideration, since the features of this type of legume are not similar 
to the others. No other legume types were considered.

2.2.2. Selecting individuals by consumption

Spot urine samples were matched to corresponding individual FFQ data. From 
a cohort of 828 individuals, 25 subjects were defined as NP consumers and 37 
as HP consumers (none of the other participants from both pulse consumer 
groups fulfilled any criteria). In order to reduce the potential sources of 
variability not related to pulse exposure, the number of HP consumers was 
balanced against NP consumers (HP = 25, NP = 25). Finally, dietary data, 
anthropometry, biochemical parameters, health status and medication were 
explored with a view to discarding any variability unrelated to pulse 
consumption.

2.3. Metabolomics analysis
2.3.1. Urine sample analysis and data processing

Morning fasting spot urine samples were collected, aliquoted, encoded and frozen at -80 °C until were use. Sample preparation was based on the methodology previously published. The 1H-NMR urinary spectra were acquired using a Varian-Inova-500 MHz NMR spectrometer with presaturation of the water resonance using a NOESYPRESAT pulse sequence. During the acquisition, the internal temperature was kept constant at 298 K. An exponential window function was applied to the free induction decay (FID) with a line-broadening factor of 0.3 Hz prior to Fourier transformation. For each sample, a total of 128 scans were collected into 32 K data points with a spectral width of 14 ppm at 300 K, an acquisition time of 3.2 s and a relaxation delay of 3 s.

1H-NMR spectra were phased, baseline-corrected and calibrated (TSP, 0.0 ppm) using TopSpin software (version 3.0, Bruker, BioSpin, Germany). After baseline correction, original spectral data were bucketed in intelligent bucketing domains of 0.005 ppm with ACD/NMR Processor 12.0 software (Advanced Chemistry Development, Toronto, Canada). The water signal and noise regions above 9.5 ppm and below 0.5 ppm were excluded from the analysis.

Data were submitted to MetaboAnalyst 3.0 for interquartile range filtering and normalization by the sum of the intensities of the spectra.34

2.3.2. Statistical analysis

The NMR data set was log-transformed, Pareto-scaled and posteriorly analysed in a multivariate approach using SIMCA-P+13.0 software (Umetrics, Umeå,
Sweden). Interindividual variation may confuse the effects of intervention, particularly in multivariate data of high dimensionality. Therefore, partial least squares discriminant analysis with orthogonal signal correction (OSC-PLS-DA) was used to explore the differences in metabolomes among the pulse consumption. OSC filtration was used to reduce the variability not associated with dietary classification, as has been done in other published nutrimetabolomic studies. The quality of the models was evaluated by the proportion of the variance of the response variable that is explained by the model (R²Y) and the predictive ability (Q²) parameters. Validation of the models and the evaluation of the degree of overfitting were carried out using a permutation test (n = 200), and the correlation coefficient between the original Y and the permuted Y plotted against the cumulative R² and Q² was calculated. Those NMR signals with variable importance for projection (VIP) values ≥1 in the component of the OSC-PLS-DA model were selected as being relevant for explaining the differences in metabolic profiles. These variables were further studied through the univariate Student’s t-test among HP and NP consumers to assess the statistical significances. Multiple tests were controlled by the false discovery rate (FDR). Statistical significance was considered at an FDR-adjusted p-value <0.05. Then, Cliff’s delta was chosen for estimation of the effect size and calculated for each feature.

2.3.3. Metabolite identification

Metabolite identification was performed using the Chenomx NMR Suite Professional Software package (version 8.1; Chenomx Inc., Edmonton,
Canada) and by comparing NMR spectral data to those available in databases such as the Human Metabolome Database (http://www.hmdb.ca), the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu) and the Madison Metabolomics Consortium Database (www.mmcd.nmrfam.wisc.edu), along with the existing NMR-based metabolomics literature. Further, a Pearson’s correlation test and clustering analysis with Pearson distance and Ward’s minimum variance using PermutMatrix 1.9.3.0 software were applied in order to identify the signals corresponding to the same metabolite.

2.4. Study of combined urinary biomarker model

The interaction between gender and the resulting metabolites was evaluated by a logistic regression for discarding any effect on the biomarkers. Then, these metabolites were submitted to a stepwise logistic regression analysis (IBM SPSS Statistics 20 software, SPSS, Inc., Chicago, IL, USA) to evaluate whether the combination of more than one biomarker improves the discrimination of pulse consumption. The models were constructed through a dichotomous variable of pulse consumption as dependent variable and identified metabolites as independent variables, with a p-value of <0.05 as a condition required for entering and remaining in the model. For validation of models, the analysis with a training set of 2/3 of the samples (removing 1/3 of the individuals as the validation set) was permuted 20 times. Spearman’s rank correlation coefficient was used to assess correlations between the combined models and pulse consumption.
The global performance of the models was evaluated by receiver operating characteristic (ROC) curve and estimation of the area under the curve (AUC) values. The optimum cut-off for sensitivity and specificity of the biomarkers was determined as the minimum distance to the top-left corner.39

3. Results

A flow chart of the participants allocated in the present study is presented in the Supplementary Information (Figure S1). Anthropometric measurements and biochemical analyses were performed using standardized methods.28 HP consumers showed a pulse consumption of 38.45 ± 14.68 g/day, while NP consumers reported a consumption of 3.75 ± 3.95 g/day (mean ± SD). The characteristics of participants classified by pulse consumption (Table S2) are presented in the Supplementary Information. The stratified populations were not different in terms of disease (type 2 diabetes mellitus or cardiovascular risk factors), medications or biochemical parameters, among other data. Subjects who were HP consumers showed higher amounts of both dietary fibre ($p < 0.01$) and polyunsaturated fatty acid ($p < 0.05$) intakes as a consequence of legume macronutrient composition.6,40 No significances other than pulses were found with regard to food intake.

3.1. Selection of significant biomarkers related to pulse consumption

For the analysis of the features belonging to pulse consumption in the urinary metabolome of the HP and NP consumers, an orthogonal signal correction was applied before PLS-DA analysis. The OSC-PLS-DA analysis of the two groups
resulted in a latent variable model with $R^2_Y$ and $Q^2$ values of 0.954 and 0.809, respectively, indicating that the model was able to classify each subject in the correct consumption group. The corresponding permutation tests showed negative $Q^2$ intercepts with a value of -0.164, implicating validation of the model.\textsuperscript{35} With the purpose of selecting the most discriminative urinary markers of consumption, only the statistically significant variables coming from both multi- and univariate analyses simultaneously were considered.

3.2. Identified biomarkers of habitual pulse consumption

A total of 16 compounds were identified as discriminant metabolites of pulse consumption. Metabolites and chemical shifts identified corresponding to statistical analyses are presented in Table 1. The total number of metabolites related to pulse consumption was divided into categories as follows: (i) choline metabolism: choline, dimethylglycine, trimethylamine-N-oxide (TMAO) and dimethylamine; (ii) protein-related compounds: 3-methylhistidine, methylguanidine, phenylalanine, glutamine and $n$-acetylglutamine; and (iii) energy metabolism: glucose, leucine, isovalerylglucose, and isobutyric, acetoacetic, citric and cis-aconitic acids.

3.3. Combined urinary biomarker approach

Logistic regression analysis revealed that there was no significant interaction between gender and the metabolites (p>0.05; all) shown in Table S3 (Supplementary Information). To study the improvement of the discrimination
between groups (HP and NP consumers), a conditional stepwise variable selection method, through a binary logistic regression analysis, was used on a combination of more than one discriminant metabolite. Table S4 (Supplementary Information) shows the resulting metabolites included in all 20 permuted models and the contribution to the model. Three metabolites were included in the fitted model according to the maximum AUC, which contained two protein-related metabolites (glutamine and 3-methylhistidine) and one choline-related metabolite (dimethylamine). These three metabolites correlated individually with the pulse consumption. However, the combined model exhibited the strongest correlation ($r=0.73$, $p<0.01$) with the pulse exposure, as shown in Table S5 (Supplementary Information).

The ROC curve analysis was used to evaluate the combined metabolite model and their metabolites using both training and validation sets separately. The highest AUC was for the combined metabolite model for both training (AUC = 95.6%) and validation (AUC = 94.4%) sets, including glutamine, 3-methylhistidine and dimethylamine followed by the individual metabolites 3-methylhistidine (AUC = 82.4%), glutamine (AUC = 81.6%) and dimethylamine (AUC = 75.0%), as shown in Figure 1. The equations generated from the logistic regression and the AUCs from the models with their sensitivity and specificity are shown in Table 2.
Figure 1. Receiver operating characteristic (ROC) curves of combined model (continuous line) with the area under the ROC curve and of included individual metabolites (discontinuous lines) in the training (A) and validation (B) sets.

4. Discussion

In this study, we present a panel of different urinary metabolites related to habitual pulse exposure using a $^1$H-NMR-based untargeted nutrimetabolomic approach in a free-living population. In addition, high correlations were found when the exposure was assessed as a continuous variable (defined by the combined biomarker panel).

4.1. Characterization of pulse fingerprinting in urine

4.1.1. Pulse metabolomic fingerprinting and choline metabolism

Several compounds found in the spot urine of pulse consumers are related to choline. Thus pulses, as a rich source of choline, may be the precursor of
additional metabolites that are susceptible to microbial degradation generating
new compounds.\textsuperscript{42} Therefore, the increase of several intermediates of choline
metabolism, such as choline itself, TMAO and dimethylamine, appears to be a
consequence of the microbial activity in HP consumers. In relation to this, De
Filippis and co-workers found an inverse correlation between urinary TMAO and
vegetarian diets compared with omnivore ones. However, they suggest different
food sources of carnitine and choline such as eggs, beef, pork and fish.\textsuperscript{43}
Hence, legumes from vegetarian diets should be proposed as a food choline
source. The increase of dimethylamine, which is also a downstream product of
choline, supports the microbial degradation of TMAO from choline. Furthermore,
TMAO was identified as a major source of urinary dimethylamine in humans.\textsuperscript{44}
directly related to gut microbiota metabolism.\textsuperscript{45} On the other hand, the increase
of urinary dimethylglycine may also come from the choline contained in pulses.
The enzymes choline dehydrogenase, betaine aldehyde dehydrogenase and
betaine homocysteine methyltransferase lead to dimethylglycine from choline.\textsuperscript{46}
Therefore, the results of the present study suggest a possible impact on urinary
metabolome by choline from pulses that is degraded via both (i) mammalian
pathways in which choline is converted to dimethylglycine through betaine, and
(ii) microbial metabolism in which choline is degraded to trimethylamine, TMAO
and dimethylamine. For this reason, we propose dimethylamine and
dimethylglycine in spot urine as potential candidates for biomarkers of pulse
consumption. Nevertheless, these choline-related metabolites need to be
further explored in controlled studies confirming that they are food intake
biomarkers instead of reflecting metabolic differences due to the pulse.
consumption. Figure 2 shows both proposed pathways for downstream products of choline.

**Figure 2.** Proposed pathways for choline degradation from pulses including significant metabolites in HP consumers in the present study. Image courtesy of Francisco Madrid-Gambin. Copyright 2016.

4.1.2. Pulse metabolomic fingerprinting and protein-related compounds

With regard to the increases in glutamine and the acetylated form n-acetylglutamine, several explanations may be proposed. Glutamine and n-acetylglutamine could come from dietary sources since glutamine is found in high-protein foods, such as pulses. Another explanation could be the alteration of urinary levels previously shown in this type of population, affected by pulse consumption. There was a higher excretion of 3-methylhistidine in HP consumers. This metabolite is a biomarker of meat and fish consumption, denoting a potential role as a biomarker of consumption. Interestingly, all food sources of this metabolite are also protein sources, including pulses as a
vegetable source, as highlighted in the present study. However, 3-methylhistidine is also a muscle protein breakdown that is sensitive to gender and age. Methylguanidine is derived from protein catabolism and from the breakdown of creatinine, therefore it may be related to protein from pulses.

4.1.3. Pulse metabolomic fingerprinting and energy metabolism

The signals of several usual metabolites were altered between the two groups. However, the definition as food intake biomarkers is controversial. Instead, they probably reflect metabolic differences associated with being a low and high consumer, based on the study design. Most of the biomarkers found in the present study are metabolites related to energy metabolism. The lower excretion of acetoacetic acid, glucose and tricarboxylic acid (TCA) cycle intermediates (citric and cis-aconitic acids) appears to involve a different energy modulation according to the pulse consumption. This fact is in part reinforced by changes in BCAAs and subproducts, which are involved in energy metabolism. For example, isobutyric acid is a short-chain fatty acid that is a product of BCAA catabolism of valine, which is a glucogenic BCAA metabolized via the methylmalonyl-CoA in the TCA cycle. On the other hand, acetoacetic acid is a ketone body produced in the human liver for fatty acid breakdown, which serves as a source of energy when normal glycolysis is altered. Interestingly, acetoacetic acid was shown to be increased in diabetes mellitus. Hence, we hypothesize that gluconeogenesis may be diminished in pulse consumers, supported by the urinary reduction of acetoacetic and isobutyric acids (lower fatty acid catabolism), and the reduction of TCA cycle intermediates and urinary
glucose (better use of glucose). Furthermore, it was observed that pulse consumption has a glucose-lowering role in diabetes mellitus, thereby explaining the lower plasma glucose concentration and lower urinary excretion. Figure 3 shows the resulting endogenous metabolites connected to the TCA cycle. Nevertheless, the small sample size that resulted after the stratification of the population leads to only exploratory results that should be confirmed.

The role of other findings such as increases of leucine and phenylalanine in pulse consumers is unclear. On the one hand, these habitual urinary compounds could be increased as a consequence of pulses being the source. However, another explanation of these findings could support the hypothesis above. Leucine, which is an acetoacetic acid precursor, may modulate glucose metabolism through oxidation, as well as insulin signalling and release. In addition, stimulation of glucose recycling via the glucose-alanine cycle by leucine may inhibit protein breakdown. However, alterations in urinary leucine have also been proposed for the prediction of diabetes mellitus, probably related to the perturbed energy metabolism. The origin of increased phenylalanine is also uncertain. This ketogenic amino acid can stimulate insulin and glucagon concentration, enhancing glucose homeostasis, and is also altered in an insulin-resistant state and obesity. Overall, the consumption of pulses seems to affect the energy metabolism in the studied population.
4.2. New biomarker panel to characterize habitual pulse consumption

To delimit the prediction of habitual pulse intake, comprising lentils, chickpeas and beans, a combination of more than one discriminatory metabolite had to be studied. The combination of three metabolites enhanced considerably the AUC and the confidence interval of the model in comparison with individual metabolites, as shown in Table 2. The developed model indicated that glutamine, 3-methylhistidine and dimethylamine were the strongest candidates for exposure biomarkers. It is important to note that the role of the component coming from choline metabolism suggests the importance of this metabolite as a biomarker of intake. Interestingly, metabolites displaying changes in energy metabolism were scarcely considered by the stepwise logistic regression. None of the other metabolites entered the model, probably as a result of collinearity in the evidence provided by these compounds, which may originate from the same...
metabolic pathways, giving similar biological or dietary information. Instead, two metabolites related to protein coming from pulses and one connected to microbiota choline degradation were established in the combined metabolite model, giving complementary information, showing a better discrimination (AUC > 90% in both training and validation sets) than each metabolite individually (AUC < 90% in all cases), and reinforcing the improved capacity of biomarker patterns to distinguish between different dietary exposures.

5. Conclusions

We applied an untargeted $^1$H-NMR-based metabolomic strategy to distinguish the urinary metabolome of habitual pulse consumption in a free-living population. Stepwise logistic regression analysis exhibited a useful approach to designing a combined urinary biomarker model taking into consideration the different characteristics of pulses. With regard to food metabolome, this study points to a central role of choline contained in pulses and breakdown products such as dimethylglycine, TMAO and dimethylamine. Protein-related compounds such as glutamine, 3-methylhistidine and methylguanidine were also increased in the urine of HP consumers. The combined metabolite model indicated that dimethylamine, 3-methylhistidine and glutamine were the strongest candidates for exposure prediction. In relation to energy metabolism, numerous compounds connected to the TCA cycle, including BCAAs and acetoacetic acid, were modified, denoting a substantial impact on energy metabolism modulation and on urinary glucose in this population. However, since the status of type 2 diabetes mellitus or three or more major cardiovascular risk factors in the
studied population could have a distinctive energy modulation, properly controlled interventions could confirm the findings observed in this cross-sectional study.

6. Supporting Information

Table S1 – Criteria for stratifying participants by frequency of consumption.

Table S2 – Characteristics of the study population according to pulse consumption.

Table S3 – Interaction between gender and the metabolites found in the present study.

Table S4 – Permutated models used in training/validation sets with the resulting metabolites.

Table S5 – Correlations between legume consumption and the combined model for prediction of legume exposure and considered individual metabolites.

Figure S1 – Flow chart of subjects from the PREDIMED subcohort included in the study.

7. Conflict of interest disclosure

The authors declare no competing financial interest.

8. Acknowledgements
This study is supported by Spanish National Grants from the Ministry of Economy and Competitiveness (MINECO), and co-funded by FEDER (Fondo Europeo de Desarrollo Regional): AGL2009-13906-C02-01, JPI HDHL FOODBALL Project (PCIN-2014-133-MINECO Spain), and the award of 2014SGR1566 from the Generalitat de Catalunya’s Agency AGAUR. We also thank the EU Joint Programming Initiative “A Healthy Diet for a Healthy Life” on Biomarkers BioNHFOODBALL. F.M-G. acknowledges the APIF PhD fellowship (University of Barcelona). M.U-S. would like to thank the “Ramón y Cajal” programme (RYC-2011-09677) from MINECO and the Fondo Social Europeo. EAA would like to thank CONACYT (Mexico) for the PhD fellowship.

**Abbreviations**

AUC, area under the curve; FFQ, food frequency questionnaire; FID, free induction decay; HP, habitual pulses; ISRCTN, International Standard Randomized Controlled Trial Number; KOD, potassium deuteroxide; MD, Mediterranean diet; NMR, nuclear magnetic resonance; NP, non-pulses; OSC-PLS-DA, partial least-squares discriminant analysis with orthogonal signal correction; ROC, receiver operating characteristic; TCA, tricarboxylic acid; TMAO, trimethylamine-N-oxide; TSP, 3-(trimethylsilyl)-propionate-2,2,3,3-d₄; VIP, variable importance projection.
References


Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res. 2015, 43 (W1), W251–W257.


the high fat-fed C57BL/6J mouse. *Diabetes, Obes. Metab.* 2008, 10 (10), 950–958.


Table 1. Tentative discriminant metabolites derived from the multi- and univariate analysis of $^1$H-NMR signal intensities in urine from HP consumers$^a$

<table>
<thead>
<tr>
<th>Source</th>
<th>Metabolite</th>
<th>HP vs NP</th>
<th>δ (multiplicity)</th>
<th>FDR p-value$^f$</th>
<th>Cliff’s delta$^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choline metabolism</td>
<td>Choline</td>
<td>↑</td>
<td>3.19 (s)</td>
<td>3.27 x 10$^{-2}$</td>
<td>0.475</td>
</tr>
<tr>
<td></td>
<td>Dimethylglycine</td>
<td>↑</td>
<td>2.93 (s)</td>
<td>3.81 x 10$^{-2}$</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>TMAO</td>
<td>↑</td>
<td>3.27 (s)</td>
<td>7.29 x 10$^{-3}$</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>Dimethylamine</td>
<td>↑</td>
<td>2.72 (s)</td>
<td>1.05 x 10$^{-2}$</td>
<td>0.488</td>
</tr>
<tr>
<td></td>
<td>N-acetylglutamine</td>
<td>↑</td>
<td>2.04 (s)</td>
<td>2.55 x 10$^{-2}$</td>
<td>0.706</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glutamine</td>
<td>↑</td>
<td>2.12 (m)</td>
<td>1.17 x 10$^{-6}$</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein-related compounds</td>
<td>Phenylalanine</td>
<td>↑</td>
<td>3.19 (m)</td>
<td>3.21 x 10$^{-2}$</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.98 (dd)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.32 (d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.36 (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.42 (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methylguanidine</td>
<td>↑</td>
<td>2.83 (s)</td>
<td>3.72 x 10$^{-4}$</td>
<td>0.635</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-Methylhistidine</td>
<td>↑</td>
<td>7.18 (s)</td>
<td>1.73 x 10$^{-4}$</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.92 (s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Citric acid</td>
<td>↓</td>
<td>2.55 (dd)</td>
<td>8.43 x 10$^{-5}$</td>
<td>-0.690</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.69 (dd)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cis-aconitic acid</td>
<td>↓</td>
<td>5.74 (s)</td>
<td>1.11 x 10$^{-3}$</td>
<td>-0.629</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.12 (s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glucose</td>
<td>↓</td>
<td>3.50 (m)</td>
<td>7.89 x 10$^{-5}$</td>
<td>-0.718</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.66 (d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.25 (d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy metabolism</td>
<td>Acetoacetic acid</td>
<td>↓</td>
<td>2.27 (s)</td>
<td>1.95 x 10$^{-2}$</td>
<td>-0.408</td>
</tr>
<tr>
<td></td>
<td>Isovalerylglucose</td>
<td>↑</td>
<td>0.92 (d)</td>
<td>2.83 x 10$^{-4}$</td>
<td>0.635</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.16 (d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.74 (d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leucine</td>
<td>↑</td>
<td>0.94 (t)</td>
<td>1.18 x 10$^{-3}$</td>
<td>0.626</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.70 (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.72 (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isobutyric acid</td>
<td>↓</td>
<td>1.06 (d)</td>
<td>1.29 x 10$^{-2}$</td>
<td>-0.446</td>
</tr>
</tbody>
</table>

$^a$All features have VIP values ≥1.0 in the corresponding OSC-PLS-DA model. $^b$FDR p-value of Student’s t-test with False Discovery Rate correction. $^c$Estimation of the effect size by Cliff’s delta with thresholds: |n|<0.330 "small", 0.330>|n|<0.474 "medium" and |n|>0.474 "large". TMAO, trimethylamine-N-oxide. s: singlet, d: doublet, t: triplet, dd: double doublet, m: multiplet.
Table 2. Receiver operating characteristic (ROC) curve parameters of combined models and of individual metabolites in both training and validation sets.

<table>
<thead>
<tr>
<th>Set†</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined model</td>
<td>Training</td>
<td>88.2</td>
<td>93.7</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>87.5</td>
<td>88.9</td>
</tr>
<tr>
<td>3-Methylhistidine</td>
<td>Training</td>
<td>76.5</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>87.5</td>
<td>77.8</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Training</td>
<td>76.5</td>
<td>81.2</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>87.5</td>
<td>77.8</td>
</tr>
<tr>
<td>Dimethylamine</td>
<td>Training</td>
<td>82.4</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>50.0</td>
<td>66.7</td>
</tr>
</tbody>
</table>

AUC: area under the ROC curve. CI: confidence interval. †Corresponding to 2/3 of the population for the training and 1/3 for the validation set.
Figure 1. Receiver operating characteristic (ROC) curves of combined model (continuous line) with the area under the ROC curve and of included individual metabolites (discontinuous lines) in the training (A) and validation (B) sets.

Figure 2. Proposed pathways for choline degradation from pulses including significant metabolites in HP consumers in the present study.

Figure 3. Modified metabolites found in HP consumers connected to energy metabolism.
For Table of Contents Only